NORMAL REGRESSION MODEL WITH CONJUGATE PRIORS

AAEC 6564
INSTRUCTOR: KLAUS MOELTNER

Textbooks: Koop (2003), Ch.3; Koop et al. (2007), Ch.10,11; Hoff (2009), Ch. 9
`Matlab` scripts: `igPlots, mod1s2a, mod1s2b`

Model Outline

"Conjugate" refers to the property of a prior to yield, when combined with the likelihood function, a posterior that has the same density family as the prior itself. The KEY notion here is that, as a result, the full statistical form of the posterior is known. This avoids the need for computational simulation to learn about the posterior (see next chapter). In the "old" days, this was an important consideration when embarking on Bayesian analysis.

Conjugate priors (usually) also produce an analytical solution to the *marginal likelihood*. That is the primary reason we use them today, especially for Bayesian model search and model averaging, where we need to evaluate the marginal likelihood many, many times, and computing time becomes an issue.

Conjugate priors can be quite restrictive, and in few real applications will we be able to formalize our prior understanding of the world in conjugate terms. However, this particular model is very useful for pedagogical purposes, and is thus a good starting point.

The regression model. Start by writing down your econometric model, linking your outcome variable to a set of explanatory variables ("regressors"). In the case of the "Classical Linear Regression Model" (CLRM, no pun intended...), we have:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \epsilon, \quad \epsilon \sim n\left(\mathbf{0}, \sigma^2 \mathbf{I}\right) \tag{1}$$

Thus, our model parameters about which we wish to learn are $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta}' & \sigma^2\end{bmatrix}$.

The likelihood function. The *sample distribution*[1] for the CLRM is given as

$$p\left(\mathbf{y}|\boldsymbol{\theta}, \mathbf{X}\right) = (2\pi)^{-n/2}\left(\sigma^2\right)^{-n/2} exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right) \tag{2}$$

[1] Other textbook terms for the sample distribution include "sampling distribution" and "sample density". I like to reserve the term "sampling distribution" for the finite sample properties of *estimators* in classical estimation. "Sample density" is fine, though.

Strictly speaking, this is *not* the likelihood function, it is only *proportional* to it. The likelihood function is expressed in terms of the unknown parameters, *given* a fixed set of data. It can be conveniently thought of components of the sample distribution that directly relate to the parameters, multiplied by a constant of proportionality. In this case, we can write:

$$\mathfrak{L}\left(\boldsymbol{\theta}|\mathbf{y},\mathbf{X}\right) \propto exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right) \tag{3}$$

In practice, the two concepts, sample distribution and likelihood function, are often used synonymously. Also, for our mathematical derivations this distinction is irrelevant. We will follow the bulk of your Bayesian textbooks and, with abuse of the term, refer to $p\left(\mathbf{y}|\boldsymbol{\theta},\mathbf{X}\right)$ with the form given in (2) as the likelihood function.

The priors. The hallmark characteristic of this conjugate model is that the prior for $\boldsymbol{\beta}$ is specified as conditional on σ^2. It follows a multivariate normal density:

$$\begin{aligned}
&\boldsymbol{\beta}|\sigma^2 \sim n\left(\boldsymbol{\mu}_0,\sigma^2\mathbf{V}_0\right), \quad \text{or}\\
&p\left(\boldsymbol{\beta}|\sigma^2\right) = (2\pi)^{-k/2}\left|\sigma^2\mathbf{V}_0\right|^{-1/2} exp\left(-\tfrac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)'\left(\sigma^2\mathbf{V}_0\right)^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)\right)
\end{aligned} \tag{4}$$

The prior for σ^2, on the other hand, is not dependent on $\boldsymbol{\beta}$. It follows an *inverse gamma* (ig) density.

$$\begin{aligned}
&\sigma^2 \sim ig\left(\nu_0,\tau_0\right), \quad \text{or}\\
&p\left(\sigma^2\right) = \frac{\tau_0^{\nu_0}}{\Gamma\left(\nu_0\right)}\left(\sigma^2\right)^{-(\nu_0+1)} exp\left(-\frac{\tau_0}{\sigma^2}\right), \quad \text{with}\\
&E\left(\sigma^2\right) = \frac{\tau_0}{\nu_0-1}, \quad V\left(\sigma^2\right) = \frac{\tau_0^2}{\left(\nu_0-1\right)^2\left(\nu_0-2\right)}
\end{aligned} \tag{5}$$

`Matlab` script `igPlots` let's you play with the *ig* density.

Looking ahead, when combined with the likelihood, these prior settings will lead to known statistical distributions for the posterior for both $\boldsymbol{\beta}$ and σ^2. Specifically, the posterior for σ^2 will also be *ig*, and the posterior for $\boldsymbol{\beta}$ will follow a t-distribution (thus bending a bit our notion of "conjugate").

The joint posterior. Next, let' s write down the joint posterior kernel. This includes all elements of the likelihood function and the priors that are *not multiplicatively separable* from parts that include the parameters. In this case:

$$\begin{aligned}
&p\left(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}\right)\propto p\left(\boldsymbol{\beta}|\sigma^2\right)p\left(\sigma^2\right)p\left(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\mathbf{X}\right)\propto\\
&|\sigma^2\mathbf{V}_0|^{-1/2}exp\left(-\tfrac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)'\left(\sigma^2\mathbf{V}_0\right)^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)\right)*\\
&\left(\sigma^2\right)^{-(\nu_0+1)}exp\left(-\frac{\tau_0}{\sigma^2}\right)*\\
&\left(\sigma^2\right)^{-n/2}exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)
\end{aligned}\tag{6}$$

As usual, this is not a kernel we can immediately associate with a well-understood statistical density. The logical next step is to examine the *conditional* posterior densities for each parameter.

Conditional posterior for $\boldsymbol{\beta}$. We start by collecting all terms from (6) that are not multiplicatively separable from components that include $\boldsymbol{\beta}$:

$$\begin{aligned}
&p\left(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X}\right)\propto\\
&exp\left(-\tfrac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)'\left(\sigma^2\mathbf{V}_0\right)^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)\right)*\\
&exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)
\end{aligned}\tag{7}$$

Some algebra manipulations:

$$\begin{aligned}
&p\left(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X}\right)\propto\\
&exp\left(-\tfrac{1}{2}\left(\boldsymbol{\beta}'\left(\sigma^2\mathbf{V}_0\right)^{-1}\boldsymbol{\beta}-2\boldsymbol{\beta}'\left(\sigma^2\mathbf{V}_0\right)^{-1}\boldsymbol{\mu}_0\right)\right)*exp\left(\tfrac{1}{2}\boldsymbol{\mu}_0'\left(\sigma^2\mathbf{V}_0\right)^{-1}\boldsymbol{\mu}_0\right)*\\
&exp\left(-\frac{1}{2\sigma^2}\left(-2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y}+\boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}\right)\right)*exp\left(-\frac{1}{2\sigma^2}\mathbf{y}'\mathbf{y}\right)\propto\\
&exp\left(\boldsymbol{\beta}'\left(\left(\sigma^2\mathbf{V}_0\right)^{-1}\boldsymbol{\mu}_0+\tfrac{1}{\sigma^2}\mathbf{X}'\mathbf{y}\right)-\tfrac{1}{2}\boldsymbol{\beta}'\left(\left(\sigma^2\mathbf{V}_0\right)^{-1}+\tfrac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}\right)
\end{aligned}\tag{8}$$

Now consider some generic multivariate normal density for random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$. The kernel of this density can always be written as

$$p\left(\mathbf{x}\right)\propto exp\left(\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-\tfrac{1}{2}\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)\tag{9}$$

Thus, we can recognize the kernel in (8) as the kernel of a multivariate normal density with variance $\mathbf{V}_1 = \left(\left(\sigma^2\mathbf{V}_0\right)^{-1} + \frac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)^{-1}$ and mean $\boldsymbol{\mu}_1 = \mathbf{V}_1 * \left(\left(\sigma^2\mathbf{V}_0\right)^{-1}\boldsymbol{\mu}_0 + \frac{1}{\sigma^2}\mathbf{X}'\mathbf{y}\right)$, i.e we can write:

$$\begin{aligned}
&\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X} \sim n\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right), \quad \text{with} \\
&\mathbf{V}_1 = \left(\left(\sigma^2\mathbf{V}_0\right)^{-1} + \tfrac{1}{\sigma^2}\mathbf{X}'\mathbf{X}\right)^{-1} = \sigma^2\left(\mathbf{V}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}, \quad \text{and} \\
&\boldsymbol{\mu}_1 = \mathbf{V}_1 * \left(\left(\sigma^2\mathbf{V}_0\right)^{-1}\boldsymbol{\mu}_0 + \tfrac{1}{\sigma^2}\mathbf{X}'\mathbf{y}\right) = \left(\mathbf{V}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{V}_0^{-1}\boldsymbol{\mu}_0 + \mathbf{X}'\mathbf{y}\right)
\end{aligned} \tag{10}$$

Notice that when $\mathbf{V}_0$ is large, the expression for $\boldsymbol{\mu}_1$ collapses to the OLS estimator.

Conditional posterior for σ^2. We start again with the joint posterior kernel in (6) and keep all elements that are not multiplicatively separable from components that include σ^2:

$$\begin{aligned}
&p\left(\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}\right) \propto \\
&|\sigma^2\mathbf{V}_0|^{-1/2} exp\left(-\tfrac{1}{2}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_0\right)'\left(\sigma^2\mathbf{V}_0\right)^{-1}\left(\boldsymbol{\beta} - \boldsymbol{\mu}_0\right)\right) * \\
&\left(\sigma^2\right)^{-(\nu_0+1)} exp\left(\frac{-\tau_0}{\sigma^2}\right) * \\
&\left(\sigma^2\right)^{-n/2} exp\left(-\frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right)
\end{aligned} \tag{11}$$

Noting that $|\sigma^2\mathbf{V}_0| = \left(\sigma^2\right)^k |V_0|$, where k is the dimension of $\boldsymbol{\beta}$, we can write this as:

$$\begin{aligned}
&p\left(\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{X}\right) \propto \\
&\left(\sigma^2\right)^{-\left(\frac{2\nu_0+k+n}{2}+1\right)} * \\
&exp\left(-\tfrac{1}{\sigma^2}\left(\tau_0 + \tfrac{1}{2}\left(\left(\boldsymbol{\beta} - \mu_0\right)'\mathbf{V}_0^{-1}\left(\boldsymbol{\beta} - \mu_0\right) + \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right)\right)\right)
\end{aligned} \tag{12}$$

We immediately recognize this as the kernel of another *ig* density. Specifically:

$$\begin{aligned}
&\sigma^2|\boldsymbol{\beta}, \mathbf{y}, \mathbf{X} \sim ig\left(\nu_1, \tau_1\right), \quad \text{with} \\
&\nu_1 = \frac{2\nu_0 + k + n}{2}, \quad \text{and} \\
&\tau_1 = \tau_0 + \tfrac{1}{2}\left(\left(\boldsymbol{\beta} - \mu_0\right)'\mathbf{V}_0^{-1}\left(\boldsymbol{\beta} - \mu_0\right) + \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right)
\end{aligned} \tag{13}$$

We could now obtain draws from the *unconditional* (or "*marginal*") posteriors $p\left(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}\right)$ and $p\left(\sigma^2|\mathbf{y}, \mathbf{X}\right)$ via *simulation* by drawing alternately and repeatedly from these *conditional* posteriors. That's the basic idea of a *Gibbs Sampler*. However, this would not reveal the true statistical form of these marginal posteriors.

The whole point of working with conjugate priors is that we can actually derive the complete statistical form for $p(\boldsymbol{\beta}|\mathbf{y},\mathbf{X})$ and $p(\sigma^2|\mathbf{y},\mathbf{X})$. Thus, a bit more algebra and calculus work will be needed to get there.

Marginal posterior for σ^2. There are numerous ways to proceed. Perhaps the mathematically least tedious is via Bayes rule:

$$p\left(\sigma^2|\mathbf{y},\mathbf{X}\right) = \frac{p\left(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}\right)}{p\left(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X}\right)} \propto \frac{p\left(\boldsymbol{\beta}|\sigma^2\right)p\left(\sigma^2\right)p\left(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\mathbf{X}\right)}{p\left(\boldsymbol{\beta}|\sigma^2,\mathbf{y},\mathbf{X}\right)} \tag{14}$$

which just states that the marginal posterior of σ^2 equals the joint posterior of σ^2 and $\boldsymbol{\beta}$, divided by the conditional posterior of $\boldsymbol{\beta}$. The simplified kernel of the numerator is given in (6). The denominator is the conditional posterior of β, i.e. the multivariate normal density with moments given in (10). Combining these terms, with focus on parts that are relevant for σ^2, and letting $\tilde{\mathbf{V}}_1 = \left(\mathbf{V}_0^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}$ yields:

$$\begin{aligned}
&p\left(\sigma^2|\mathbf{y},\mathbf{X}\right) \propto \\
&\left(\sigma^2\right)^{-\left(\frac{2\nu_0+k+n}{2}+1\right)} * exp\left(-\tfrac{1}{\sigma^2}\left(\tau_0 + \tfrac{1}{2}\left(\left(\boldsymbol{\beta}-\mu_0\right)'\mathbf{V}_0^{-1}\left(\boldsymbol{\beta}-\mu_0\right) + \left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\right)\right) * \\
&\left(\left(\sigma^2\right)^{-k/2} exp\left(-\tfrac{1}{2\sigma^2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)\right)^{-1} = \\
&\left(\sigma^2\right)^{-\left(\frac{2\nu_0+n}{2}+1\right)} * \\
&exp\left(-\tfrac{1}{\sigma^2}\left(\tau_0 + \tfrac{1}{2}\left(\left(\boldsymbol{\beta}-\mu_0\right)'\mathbf{V}_0^{-1}\left(\boldsymbol{\beta}-\mu_0\right) + \left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) - \left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)\right)\right)
\end{aligned} \tag{15}$$

We're not done, since we still have $\boldsymbol{\beta}$ on the right hand side. However, it can be shown that:

$$\begin{aligned}
&\left(\boldsymbol{\beta}-\mu_0\right)'\mathbf{V}_0^{-1}\left(\boldsymbol{\beta}-\mu_0\right) + \left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) - \left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right) = \\
&SSE + \left(\mathbf{b}-\boldsymbol{\mu}_0\right)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}\left(\mathbf{b}-\boldsymbol{\mu}_0\right), \quad \text{where} \\
&\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}, \quad \text{and} \\
&SSE = \left(\mathbf{y}-\mathbf{X}\mathbf{b}\right)'\left(\mathbf{y}-\mathbf{X}\mathbf{b}\right)
\end{aligned} \tag{16}$$

Thus, we can write the marginal posterior kernel for σ^2 as

$$\begin{aligned}
&p\left(\sigma^2|\mathbf{y},\mathbf{X}\right) \propto \\
&\left(\sigma^2\right)^{-\left(\frac{2\nu_0+n}{2}+1\right)} * \\
&exp\left(-\tfrac{1}{\sigma^2}\left(\tau_0 + \tfrac{1}{2}\left(SSE + \left(\mathbf{b}-\boldsymbol{\mu}_0\right)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}\left(\mathbf{b}-\boldsymbol{\mu}_0\right)\right)\right)\right)
\end{aligned} \tag{17}$$

We recognize this as another *ig* density. Thus:

$$\begin{aligned}
&\sigma^2|\mathbf{y},\mathbf{X} \sim ig\left(\nu_1,\tau_1\right), \quad \text{with} \\
&\nu_1 = \frac{2\nu_0+n}{2}, \quad \text{and} \\
&\tau_1 = \tau_0 + \tfrac{1}{2}\left(SSE + \left(\mathbf{b}-\boldsymbol{\mu}_0\right)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}\left(\mathbf{b}-\boldsymbol{\mu}_0\right)\right)
\end{aligned} \tag{18}$$

Marginal posterior for $\boldsymbol{\beta}$. This density is best derived by integrating out σ^2 from the joint posterior. Specifically:

$$p\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}\right) = \int p\left(\boldsymbol{\beta},\sigma^2|\mathbf{y},\mathbf{X}\right)d\sigma^2 = \int p\left(\boldsymbol{\beta}|\sigma^2\mathbf{y},\mathbf{X}\right)p\left(\sigma^2|\mathbf{y},\mathbf{X}\right)d\sigma^2 \tag{19}$$

Again, we have everything we need: The first expression in the integral is the conditional posterior of $\boldsymbol{\beta}$ from (10), and the second term is the marginal posterior for σ^2 from (18). We can ignore all components of these two densities that do not include $\boldsymbol{\beta}$, and write:

$$\begin{aligned}
&p\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}\right) \propto \\
&\int \left(\sigma^2\right)^{-\left(\nu_1+\frac{k}{2}+1\right)} * exp\left(-\tfrac{1}{\sigma^2}\left(\tau_1 + \tfrac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)\right)d\sigma^2
\end{aligned} \tag{20}$$

Now it's time for the "integration trick": We recognize the integrand as another *ig* kernel with shape $\nu_2 = \nu_1 + \frac{k}{2}$ and scale $\tau_2 = \tau_1 + \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)$. We also know that integrating the *full* density over the entire support of a random variable has to yield "1", i.e.

$$\int \frac{\tau_2^{\nu_2}}{\Gamma\left(\nu_2\right)} * \left(\sigma^2\right)^{-\left(\nu_2+1\right)} exp\left(-\tfrac{\tau_2}{\sigma^2}\right)d\sigma^2 = 1 \tag{21}$$

Thus, we can solve the integral in (20) as:

$$\int \left(\sigma^2\right)^{-\left(\nu_2+1\right)} exp\left(-\tfrac{\tau_2}{\sigma^2}\right)d\sigma^2 = \frac{\Gamma\left(\nu_2\right)}{\tau_2^{\nu_2}} \tag{22}$$

Thus, we have

$$
\begin{aligned}
&p\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}\right)\propto\\
&\frac{\Gamma\left(\nu_2\right)}{\tau_2^{\nu_2}}\propto\tau_2^{-\nu_2}=\\
&\left(\tau_1+\tfrac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)^{-\left(\nu_1+\frac{k}{2}\right)}=\\
&\left(\tfrac{\tau_1}{2\nu_1}\left(2\nu_1+\tfrac{\nu_1}{\tau_1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\tilde{\mathbf{V}}_1^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)\right)^{-\left(\nu_1+\frac{k}{2}\right)}\propto\\
&\left(2\nu_1+\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\left(\tfrac{\tau_1}{\nu_1}\tilde{\mathbf{V}}_1\right)^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)^{-\left(\nu_1+\frac{k}{2}\right)}=\\
&\left(2\nu_1\left(1+\tfrac{1}{2\nu_1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\left(\tfrac{\tau_1}{\nu_1}\tilde{\mathbf{V}}_1\right)^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)\right)^{-\left(\nu_1+\frac{k}{2}\right)}\propto\\
&\left(1+\tfrac{1}{2\nu_1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)'\left(\tfrac{\tau_1}{\nu_1}\tilde{\mathbf{V}}_1\right)^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_1\right)\right)^{-\left(\nu_1+\frac{k}{2}\right)}
\end{aligned}
\tag{23}
$$

Now consider the generic multivariate t-density for some k-dimensioned random vector $\boldsymbol{\theta}$ with *mean* μ, *scale matrix* $\boldsymbol{\Sigma}$, and *degrees of freedom* ν. It is given as (using again the parameterization in Gelman et al. (2004), p.576):

$$
p\left(\boldsymbol{\theta}\right)=\frac{\Gamma\left(\frac{\nu}{2}+\frac{k}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\nu^{\frac{k}{2}}\pi^{\frac{k}{2}}}\left|\boldsymbol{\Sigma}\right|^{-\frac{1}{2}}*\left(1+\tfrac{1}{\nu}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\theta}-\boldsymbol{\mu}\right)\right)^{-\left(\frac{\nu}{2}+\frac{k}{2}\right)}
\tag{24}
$$

We can now recognize the marginal posterior of $\boldsymbol{\beta}$ as the kernel of a multivariate t density. Specifically:

$$
\begin{aligned}
&\boldsymbol{\beta}|\mathbf{y},\mathbf{X}\sim t\left(\mu_2,\mathbf{V}_2,\nu_2\right)\quad with\\
&\mu_2=\mu_1=\left(\mathbf{V}_0^{-1}+\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{V}_0^{-1}\boldsymbol{\mu}_0+\mathbf{X}'\mathbf{y}\right),\\
&\mathbf{V}_2=\tfrac{\tau_1}{\nu_1}\tilde{\mathbf{V}}_1=\tfrac{\tau_1}{\nu_1}\left(\mathbf{V}_0^{-1}+\mathbf{X}'\mathbf{X}\right)^{-1}\\
&\nu_2=2\nu_1=2\nu_0+n
\end{aligned}
\tag{25}
$$

Marginal Likelihood. As mentioned above, an extra "bonus" of this model is the existence of an analytical ("closed-form") solution to the marginal likelihood, $p\left(\mathbf{y}\right)$. This is extremely useful for model comparison. Note that the marginal likelihood for our observed data will not be a recognizable distribution. Thus, we can't invoke "proportionality" and aim for some kernel. We need to derive it *exactly*. Formally:

$$
\begin{aligned}
&p(\mathbf{y}|X) = \int_{\sigma^2}\int_{\boldsymbol{\beta}} p\left(\mathbf{y}, \boldsymbol{\beta}, \sigma^2 | \mathbf{X}\right) d\boldsymbol{\beta} d\sigma^2 = \\
&\int_{\sigma^2}\int_{\boldsymbol{\beta}} p\left(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X}\right) p\left(\boldsymbol{\beta}|\sigma^2\right) p\left(\sigma^2\right) d\boldsymbol{\beta} d\sigma^2 = \\
&\int_{\sigma^2}\left(\int_{\boldsymbol{\beta}} p\left(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X}\right) p\left(\boldsymbol{\beta}|\sigma^2\right) d\boldsymbol{\beta}\right) p\left(\sigma^2\right) d\sigma^2
\end{aligned}
\tag{26}
$$

The integral over $\boldsymbol{\beta}$ is taken over the likelihood in (2) times the (conditional) prior of $\boldsymbol{\beta}$ from (4). We can write this as

$$
\begin{aligned}
&\int_{\boldsymbol{\beta}} p\left(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X}\right) p\left(\boldsymbol{\beta}|\sigma^2\right) d\boldsymbol{\beta} = \\
&(2\pi)^{-\frac{n}{2}}\left(\sigma^2\right)^{-\frac{n}{2}}(2\pi)^{-\frac{k}{2}}\left(\sigma^2\right)^{-\frac{k}{2}}|V_0|^{-\frac{1}{2}} * \\
&\int exp\left(-\tfrac{1}{2\sigma^2}\left((\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + (\boldsymbol{\beta}-\boldsymbol{\mu}_0)'\mathbf{V}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_0)\right)\right) d\boldsymbol{\beta}
\end{aligned}
\tag{27}
$$

Using the result from (16), this can be expressed as

$$
\begin{aligned}
&\int_{\boldsymbol{\beta}} p\left(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \mathbf{X}\right) p\left(\boldsymbol{\beta}|\sigma^2\right) d\boldsymbol{\beta} = \\
&(2\pi)^{-\frac{n}{2}}\left(\sigma^2\right)^{-\frac{n}{2}}(2\pi)^{-\frac{k}{2}}\left(\sigma^2\right)^{-\frac{k}{2}}|V_0|^{-\frac{1}{2}} * \\
&exp\left(-\tfrac{1}{2\sigma^2}\left(SSE + (\mathbf{b}-\boldsymbol{\mu}_0)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}(\mathbf{b}-\boldsymbol{\mu}_0)\right)\right) * \\
&\int exp\left(-\tfrac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu}_1)'\mathbf{V}_1^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu}_1)\right) d\boldsymbol{\beta}
\end{aligned}
\tag{28}
$$

We recognize the integrand in the last line of (28) as the kernel of the multivariate normal. Using again our "integration trick", this has to solve to the inverse of the normalizing constant for the multivariate normal. Thus:

$$
\begin{aligned}
&\int_{\boldsymbol{\beta}} p\left(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\mathbf{X}\right) p\left(\boldsymbol{\beta}|\sigma^2\right) d\boldsymbol{\beta} = \\
&(2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{n}{2}} (2\pi)^{-\frac{k}{2}} \left(\sigma^2\right)^{-\frac{k}{2}} |V_0|^{-\frac{1}{2}} * \\
&exp\left(-\tfrac{1}{2\sigma^2}\left(SSE + (\mathbf{b}-\boldsymbol{\mu}_0)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}(\mathbf{b}-\boldsymbol{\mu}_0)\right)\right) * \\
&(2\pi)^{\frac{k}{2}} \left(\sigma^2\right)^{\frac{k}{2}} |\tilde{V}_1|^{\frac{1}{2}} = \\
&(2\pi)^{-\frac{n}{2}} \left(\sigma^2\right)^{-\frac{n}{2}} |V_0|^{-\frac{1}{2}} |\tilde{V}_1|^{\frac{1}{2}} * \\
&exp\left(-\tfrac{1}{2\sigma^2}\left(SSE + (\mathbf{b}-\boldsymbol{\mu}_0)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}(\mathbf{b}-\boldsymbol{\mu}_0)\right)\right)
\end{aligned}
\tag{29}
$$

Continuing from (26) we can multiply this by the prior for σ^2 and solve for the integral over σ^2:

$$
\begin{aligned}
&p\left(\mathbf{y}|\mathbf{X}\right) = \\
&(2\pi)^{-\frac{n}{2}} |V_0|^{-\frac{1}{2}} |\tilde{V}_1|^{\frac{1}{2}} \frac{\tau_0^{\nu_0}}{\Gamma\left(\nu_0\right)} * \\
&\int \left(\sigma^2\right)^{-\left(\frac{2\nu_0+n}{2}+1\right)} exp\left(-\tfrac{1}{\sigma^2}\left(\tau_0 + \tfrac{1}{2}\left(SSE + (\mathbf{b}-\boldsymbol{\mu}_0)'\left(\mathbf{V}_0 + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)^{-1}(\mathbf{b}-\boldsymbol{\mu}_0)\right)\right)\right) d\sigma^2
\end{aligned}
\tag{30}
$$

We recognize the integrand as the kernel of the marginal posterior for σ^2 from (17). It has to solve to the inverse of the normalizing constant. This produces the final result:

$$
p\left(\mathbf{y}|\mathbf{X}\right) = \frac{\tau_0^{\nu_0}\Gamma\left(\nu_1\right)}{(2\pi)^{\frac{n}{2}}\Gamma\left(\nu_0\right)} |V_0|^{-\frac{1}{2}} |\tilde{V}_1|^{\frac{1}{2}} \tau_1^{-\nu_1} \tag{31}
$$

where ν_1 and τ_1 are the shape and scale of the marginal posterior of σ^2 from (18).

References

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*, 2nd edn, Chapman & Hall/CRC.
Hoff, P. (2009). *A first course in Bayesian statistical methods*, Springer.
Koop, G. (2003). *Bayesian Econometrics*, Wiley.
Koop, G., Poirier, D. and Tobias, J. (2007). *Bayesian Econometric Methods*, Cambridge University Press.